Abstract

System and apparatus for using block-level sampling for histograms construction as well as distinct-value estimations. For histogram construction, the system implements

5   a two-phase adaptive method in which the sample size required to reach a desired accuracy is decided based on a first phase sample. This method is significantly faster than previous iterative block-level sampling methods proposed for the same problem. For distinct-value estimation, it is shown that existing estimators designed for uniform-random samples may perform very poorly with block-level samples. An exemplary

10   system computes an appropriate subset of a block-level sample that is suitable for use with most existing estimators.